STATWAY™ INSTRUCTOR NOTES

Lesson 3.2.4
Special Properties of the Least Squares Regression Line (Optional)

INSTRUCTOR SPECIFIC MATERIAL IS INDENTED AND APPEARS IN GREY

ESTIMATED TIME

50 minutes

MATERIALS REQUIRED

Calculator

BRIEF DESCRIPTION

This lesson is designed to help the students understand how to use summary statistics to find the LSR line.

LEARNING GOALS

Students will understand that:

- The least squares regression lines contains the point $(\bar{x},\ \bar{y})$
- The slope of the least squares regression line is related to the correlation in that a change of one standard deviation in *x* produces a fractional change of *r* standard deviations in *y*.
- The equation of the least squares line can be generated from summary statistics for *x* and *y*. The slope of the least squares line is $\frac{r \cdot s_y}{s_x}$ and its *y*-intercept is $\bar{y} - slope \cdot \bar{x}$

Lesson 3.2.4
Special Properties of the Least Squares Regression Line (Optional)

Students will be able to:

- Use the fact that the slope of the least squares line is $\frac{r \cdot s_y}{s_x}$ and its y-intercept is $\overline{y} - slope \cdot \overline{x}$ to generate the equation of the least squares line from summary statistics.

INTRODUCTION

This lesson requires students to understand that if a point is on a line, its *x*- and *y*-coordinates satisfy the relationship defined by the line. This idea is used implicitly in the investigations laid out in this lesson. If students are having trouble following the investigation, this may be at the heart of their difficulty. Therefore, be ready to discuss this issue if you diagnose that students do not understand this.

Note: This lesson is designed as an interactive lecture. Your goal is to develop the following special properties of the least squares line:

- The least squares line contains the point $(\overline{x},\ \overline{y})$. This means that when $x = \overline{x}$, the least squares line predicts.
- When *x* increases from the mean of *x* by one standard deviation in *x*, the least squares line predicts that *y* increases by *r* standard deviations in *y*. This means that the least squares line has a slope of $\frac{r \cdot s_y}{s_x}$.
- The initial value of the least squares line is $\overline{y} - \frac{r \cdot s_y}{s_x} \cdot \overline{x}$

This interactive lecture investigates three questions. The investigation of each question is broken down into smaller questions that you can pose to students. Remember to build in think time. Students will be filling in a discussion outline as you lecture, so you also need to build in time for students to take notes. Consider allowing time periodically for students to compare notes with a classmate, summarize, formulate questions, and so on.

Lesson 3.2.4
Special Properties of the Least Squares Regression Line (Optional)

STUDENT MATERIAL IS NOT INDENTED AND APPEARS IN BLACK

INTRODUCTION

In this lesson, you are interested in understanding special properties of the least squares line. You will investigate and answer the following three questions about the least squares line in this lesson:

1 What does the least squares line predict for *y* when *x* is equal to the mean of the *x* values in the sample? Here is the same question written with symbols: When $x = \overline{x}$, $\hat{y} = ?$

2 What does the least squares line predict for *y* when *x* is one standard deviation above the mean. Here is the same question written with symbols: When $x = \overline{x} + s_x, \hat{y} = ?$

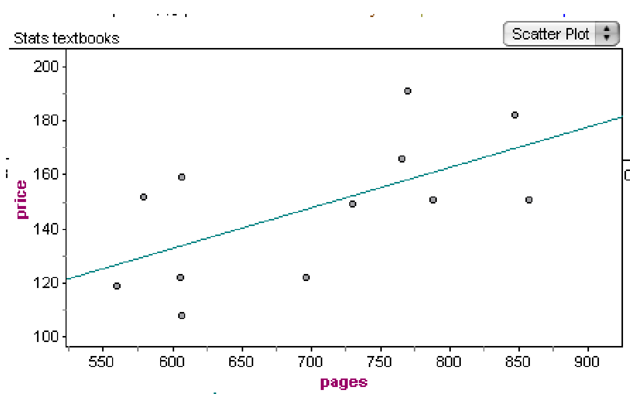
3 What is the connection between the equation of the least squares line and the summary statistics for x and y, such as means, standard deviations, and the correlation coefficient?

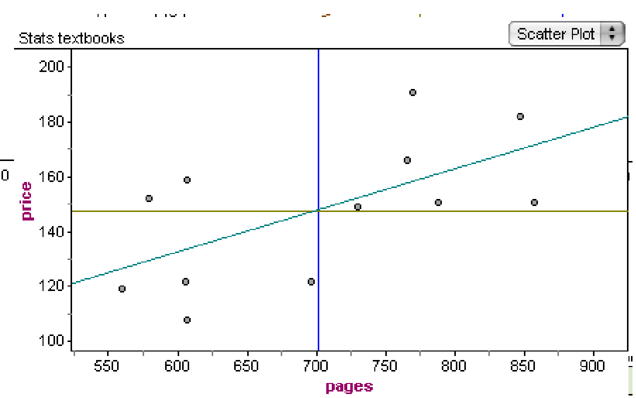
You will investigate these three questions using data on 12 statistics textbooks.

Here you have the summary statistics describing the number of pages and the publisher's list price for 12 popular statistics textbooks.

	Mean	Standard Deviation
Pages	$\overline{x} = 701.083$	$s_x = 106.402$
Price	$\overline{y} = 147.608$	$s_y = 25.718$

Below on the left, there is a scatterplot of the data and the least squares line $\hat{y} = 42.16 + 0.1504(pages)$. The correlation coefficient is 0.62. On the right, there is the same graph with the addition of the two mean lines.





Lesson 3.2.4
Special Properties of the Least Squares Regression Line (Optional)

Investigation and Discussion of Question 1

What does the least squares line predict for y when x is the mean? Here is the same question written with symbols: When $= \overline{x},\ \hat{y} = ?$

Can you figure this out from the given information before doing any calculations?

Note: Give students a minute to think about this before calling a student to answer the question or answering the question yourself.

Answer: The least squares line goes through $(\overline{x},\ \overline{y})$. This means that when $x = \overline{x}$, $\hat{y} = \overline{y}$. Tell students that this is always true for least squares lines.

If you substitute $x\ =\ 701.083$ into the given equation of the least squares line, what is the predicted value for the price? Can you figure this before you do any calculations?

Answer: This is really a check for understanding. You should get the mean of y, 147.608. You get $\hat{y} = 42.16 + 0.1504(701.083) = 147.603$ when you do the calculation; there is a bit of error due to rounding.

Investigation and Discussion of Question 2

What does the least squares line predict for y when x is one standard deviation above the mean? Here is the same question written with symbols: When $x = \overline{x} + s_x, \hat{y} = ?$

Note: This is a sophisticated question, though students might volunteer the conjecture that $\hat{y} = \overline{y} + s_y$. If they do not, suggest this as a reasonable conjecture to investigate. Then pose the question, "How could you test your conjecture?" Give students a few minutes to think about this. Then proceed with the guided investigation. Alternatively, you could call on students for ideas and build the investigation from student responses.

To answer this question, let's start by determining whether y is one standard deviation above the mean for y when x is one standard deviation above the mean for x. Here is this idea written in symbols: When $x = \overline{x} + s_x$, does $\hat{y} = \overline{y} + s_y$?

The expression $x = \overline{x} + s_x$ tells you that you are imagining looking at a textbook where the number of pages is one standard deviation above the mean number of pages for the books in the sample. How many pages is this?

Answer: $x = \overline{x} + s_x$ = 701.083 + 106.402 = 807.485, approximately 807.

Lesson 3.2.4
Special Properties of the Least Squares Regression Line (Optional)

What is the predicted price of a textbook with this many pages?

Answer: $\hat{y} = 42.16 + 0.1504(807) \approx 163.5$

Let's see if this prediction is one standard deviation in price above the mean price for the books in the sample.

Answer: $\overline{y} + s_y$ = 147.608 + 25.718 = 173.326, approximately 173.

So is your conjecture right?

Answer: No, the predicted price is about $10 cheaper than one standard deviation above the mean price.

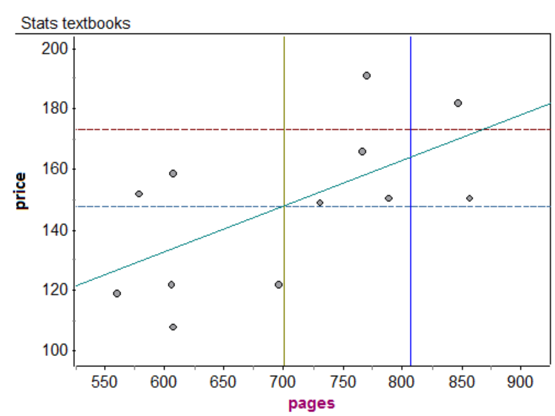
Let's summarize what is known so far. When the number of pages is one standard deviation above the mean number of pages, the predicted price increases (*choose one:* less than, more than) one standard deviation above the mean price.
Let's continue to investigate Question 2 and see if you can be more specific. Remember that you are trying to determine the predicted value when $x = \overline{x} + s_x$. You will try to answer the following questions to be more specific:

- What fraction of a standard deviation in price does the price increase? Half of a standard deviation? A third of a standard deviation? Some other fractional part of a standard deviation?
- Can you give an answer that is true for all least squares lines?

Label the following on the graph:

- the least squares line,
- the point $(\overline{x},\ \overline{y})$,
- the point $(\overline{x} + s_x,\ \overline{y} + s_y)$, and
- the point $(\overline{x} + s_x,\ \hat{y})$.

Hint: The calculations you did previously can help you label these points.



Lesson 3.2.4
Special Properties of the Least Squares Regression Line (Optional)

Note: Emphasize how the graph shows that $\hat{y}$ for $\bar{x} + s_x$ is less than $\bar{y} + s_y$. Show this on the slope triangle connecting $(\bar{x},\ \bar{y})$ and $(\bar{x} + s_x,\ \hat{y})$. Show how the change in predicted price, represented by a vertical distance on the slope triangle, is less than a standard deviation in price, represented by a vertical distance longer than the side of the slope triangle. Refer to the contextualized meaning of these symbols. For example, you know that when pages increase by one standard deviation from the mean number of pages, the predicted price increases less than a standard deviation from the mean price.

Is the increase in predicted price (*choose one:* more than or less than) *half* a standard deviation in price, $0.5s_y$? How can you tell?

Answer: You can see that the predicted price is a little more than half a price standard deviation. Show students that the vertical distance of the slope triangle, is a little more than half the vertical distance between $\bar{y}$ and $\bar{y} + s_y$. You can also calculate the increase in predicted price, $\hat{y} - \bar{y}$ = 163.5 – 147.6 = 15.9, and compare it to $0.5s_y$, 0.5(25.7) = 12.85.

Which of the expressions below best describes the increase in the predicted price? How can you tell?

- about $0.3s_y$
- about $0.6s_y$
- about $0.9s_y$

Answer: About $0.6s_y$. Explain this using the same geometric estimate as outlined just above. You can also calculate the options and compare to 15.9.

Now let's actually calculate what fraction of a standard deviation the price increases. The standard deviation in price is _____ and the increase in the predicted price from the mean line is $\hat{y} - \bar{y} =$ __________ So what fraction of a price standard deviation did $\hat{y}$ increase from $\bar{y}$?

Answers: The standard deviation in price is 25.7; increase from mean line is 15.9; 15.9/25.7 = 0.62.
Does this number have significance when you look back at the descriptive statistics for the data?

Answer: This is the value of the correlation coefficient, *r. Wow.* This will always be true for the least squares line.

Lesson 3.2.4

Special Properties of the Least Squares Regression Line (Optional)

Note: Now go back and highlight the general measurements for the slope triangle connecting $(\overline{x},\ \overline{y})$ and $(\overline{x}+s_x,\ \hat{y})$: horizontal measurement is s_x, the vertical measurement is $r \cdot s_y$.

So now you have an answer to the second question: *What does the least squares line predict for y when x is one standard deviation above the mean? When* $x = \overline{x} + s_x$, $\hat{y} =$ _____.

Answer: $\hat{y} = \overline{y} + r \cdot s_y$

Investigation and Discussion of Question 3

What is the connection between the equation of the least squares line and the summary statistics for x and y, such as means, standard deviations, and the correlation coefficient? You know from your previous work that the equation of a line has the form

$y = initial\ value + slope \cdot x$. From the slope triangle you have drawn, you can derive an important characteristic about the slope of the least squares line. The slope is the change in y divided by the change in x.

How can you determine the slope of the least squares line using summary statistics?

Answer: Use the idea that the slope is the ratio of the change in y to the corresponding change in x, so the slope of the least squares line is $\frac{r \cdot s_y}{s_x}$. Alternatively, if your students are familiar with the formula for slope, you could find the slope using the points that they now know are on the line $(\overline{x},\ \overline{y})$and $(\overline{x}+s_x,\ \overline{y}+s_y)$. However, this may easier to communicate using a slope triangle than manipulating the algebraic expressions.

You also know that the least squares line contains the point $(\overline{x}, \overline{y})$. So, you have enough information to write an expression for the y-intercept of the least squares line in terms of summary statistics.

Note: Demonstrate how the initial value can be written in terms of summary statistics. This line of reasoning involves a bit of algebra, so some students may have difficulty following the demonstration. Ultimately, you just want students to understand that the equation of the least squares line can be derived from summary statistics even if they cannot reproduce the development of the formula for initial value.)

Lesson 3.2.4
Special Properties of the Least Squares Regression Line (Optional)

Demonstration of the Development of the Formula for Initial Value

For any line $y = initial\ value + slope \cdot x$

The least squares line contains $(\bar{x}, \bar{y})$, so for the least squares line you can write $\overline{y} = y - intercept + slope \cdot \overline{x}$

You can rewrite this equation as a formula for initial value:

y-intercept $= \overline{y} - slope \cdot \overline{x}$

Rewriting the slope formula for the least squares line, you have the initial value expressed in terms of summary statistics:

y-intercept $= \overline{y} - \frac{r \cdot s_y}{s_x} \cdot \overline{x}$

Lesson 3.2.4

Special Properties of the Least Squares Regression Line (Optional)

TAKE IT HOME

1 At Los Medanos College, a statistics instructor posted the following information on her office door at the end of the semester:

Statistics FA 2010	**Mean**	**Standard Deviation**	**Correlation**
Pre-final exam average	75	8	0.7
Final exam score	78	12	

A Final course grades have not been posted. So Karen wants to predict her final exam score based on this information. She has an 82 pre-final exam average. What does the least squares line predict for Karen's final exam score?

> **Answer:** slope of regression line is $\frac{r \cdot s_y}{s_x} = 1.05$.
> y-intercept of the regression line is y-intercept $= \overline{y} - \frac{r \cdot s_y}{s_x} \cdot \overline{x} = -0.75$.
> Karen's predicted final exam score is 85.35.

B What statement is the most accurate advice you could give Karen?

- Before using the least squares line to predict your final exam score, you need to know that the relationship between pre-final exam average and final exam scores is linear. You are not given enough information to determine if the relationship is linear. Therefore, proceed with caution in using a line to make predictions in this situation.
- The correlation coefficient is 0.7. This tells you that the relationship between pre-final exam average and final exam score is fairly linear. However, there will be some error in the prediction, so do not be surprised if your final exam score differs from your prediction.

> **Answer:** The first statement is the best advice. Always look at the data to make sure a linear model is appropriate. Lesson 3.1.3 demonstrates that r-values alone do not guarantee that the data are linear.

2 In this lesson, you learned that for any least squares regression line if $= \overline{x}$, then $\hat{y} = \overline{y}$. Explain in words what this means.

> **Answer:** Start with the mean of the data's x-values. Plug this into the regression equation. The result is the mean of the data's y-values. … Of course, students may have a less polished way of communicating this idea ☺.

Lesson 3.2.4
Special Properties of the Least Squares Regression Line (Optional)

3 You conjectured that for the least squares line if $x = \overline{x} + s_x$, then $\hat{y} = \overline{y} + s_y$.

A Explain in words what this means.

> **Answer:** This conjecture says that if x increases one standard deviation above its mean, then the least squares regression line predicts that the same is true for y. In other words, y is predicted to increase by one standard deviation from its mean.

B This conjecture was not true for the statistics textbook data. Would this conjecture ever be true? If so, describe the relationship you would see in the data. If not, explain why this will never be true for any least squares line.

> **Answer:** The conjecture is only true when $r = 1$. The line has to predict an increase of one standard deviation in y for one standard deviation in x. Students saw in class that when x increases one standard deviation, y increases a fraction of a standard deviation. The change in y is $r \cdot s_y$. So r has to be 1 if y increases a full standard deviation.

+++++

This lesson is part of STATWAY™, A Pathway Through College Statistics, which is a product of a Carnegie Networked Improvement Community that seeks to advance student success. Version 1.0, A Pathway Through Statistics, Statway™ was created by the Charles A. Dana Center at the University of Texas at Austin under sponsorship of the Carnegie Foundation for the Advancement of Teaching. This version 1.5 and all subsequent versions, result from the continuous improvement efforts of the Carnegie Networked Improvement Community. The network brings together community college faculty and staff, designers, researchers and developers. It is an open-resource research and development community that seeks to harvest the wisdom of its diverse participants in systematic and disciplined inquiries to improve developmental mathematics instruction. For more information on the Statway Networked Improvement Community, please visit **carnegiefoundation.org.** For the most recent version of instructional materials, visit **Statway.org/kernel.**

+++++